Experiences in Deploying Test Arenas for Autonomous Mobile Robots

Adam Jacoff, Elena Messina, John Evans
Intelligent Systems Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

ABSTRACT

The National Institute of Standards and Technology has created a set of reference test arenas for evaluating the performance of mobile autonomous robots performing urban search and rescue tasks. The arenas are intended to help accelerate the robotic research community's advancement of mobile robot capabilities. The arenas have been deployed in two competitions thus far and are also being used by researchers to test their systems' capabilities. We describe the arenas, their use in competitions and our near-term and long-term plans for the arenas.

1. INTRODUCTION

The National Institute of Standards and Technology (NIST) has been collaborating with other government agencies and university researchers to develop methods of evaluating and measuring the performance of robotic and other intelligent systems. The community agrees that it would benefit from having uniform, reproducible means of measuring capabilities of their systems to evaluate which approaches are superior under which circumstances, and to help communicate results. One of the efforts in the performance metrics program at NIST is the creation of reference test arenas for autonomous mobile robots. The first set of arenas was modeled after the Urban Search and Rescue (USAR) application and was designed to represent, at varying degrees of verisimilitude, challenges associated with collapsed structures. This is a domain that is very dangerous for rescue personnel and in which robots will likely be able to provide increasing levels of assistance in searching for survivors. [1] The arenas were first deployed at the American Association for Artificial Intelligence (AAAI) Rescue Robot Competition in 2000. In 2001, the arenas were used at the International Joint Conference on Artificial Intelligence (IJCAI). They will again be used at AAAI-2002. Additionally for 2002 and henceforth, the RoboCup Federation [3] will use the arenas to host their newly formed RoboCupRescue league competitions. A discussion of the details of these competitions is contained in Section 3 of this paper.

There are three sets of customers for the arenas. The first are researchers, who need testing opportunities. The repeatable obstacles (sensory and physical) that are focussed towards mobile robotic perception and intelligent behavior provide them with challenges for their robots. The second are the sponsors of research. They can use the arenas for validation exercises to objectively evaluate robots in structured, repeatable, representative environments. The arenas can be used to validate robotic purchases, identify strengths and weaknesses in systems, and compare the cost effectiveness of different approaches. Finally, the end users of the robots can benefit from the resulting performance metrics. The eventual goal is to develop standard performance metrics from the arenas that can be used by purchasers to evaluate mobile robot capabilities.

There were several motivating factors for building the arenas. The first was the desire to be able to compare "apples to apples" in a technological sense. When researchers publish results, they typically describe the performance of their systems in their laboratory or demonstration environments, making it difficult to compare and contrast with others researchers' results. Isolating tests for sensing, behaviors, and other robotic capabilities – and making these tests reproducible – allows the research community to make meaningful comparisons of algorithms, sensors, platforms, and other independent items. A standardization of these challenges, through use of the arenas, enables a direct comparison of approaches.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2001**	2. REPORT TYPE	3. DATES COVERED **00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE **Experiences in Deploying Test Arenas for Autonomous Mobile Robots**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **National Institute of Standards and Technology (NIST),Manufacturing Engineering Laboratory,100 Bureau Dr,Gaithersburg,MD,20899**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the 2001 Performance Metrics for Intelligent Systems Workshop (PerMIS ?01), Mexico City, Mexico on 4 Sep 2001

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

A second desire was being able to "teach to the test." The arenas provide an objective set of measures for evaluating different robotic implementations. The arenas are not idealized "blocks world" tests. They provide some fairly realistic challenges that mobile robots must be able to address to be considered capable in this domain. We hasten to add that the USAR domain is extremely challenging. Although the arenas do provide some elements of what may be encountered in a collapsed building, they are not representative of the reality of a disaster scene. Rather, they provide a step-wise abstraction of such challenges in an attempt to isolate and repeatably test specific robot capabilities.

Another concern of research sponsors and of researchers themselves is the slowing of progress due to re-invention of the wheel. When building a robot, numerous hardware and software subsystems are required and it is not possible (or very difficult) to reuse any work done by other organizations. By highlighting successful approaches negotiating well-known obstacles, it is hoped that others will better understand and adopt these approaches, and expedite their progress into other areas of research.

Finally, practice makes perfect: arenas that are available to researchers year-round should enable them to repeat experiments and therefore debug and improve their systems. The arenas are set up near the NIST campus in Gaithersburg, Maryland, and can be used by researchers year-round. Since robustness comes through repetition and testing outside perceived limits, the three arenas provide increasing levels of difficulty, so that researchers can move on to new challenges once they master the simpler sections.

2. DESIGN CONSIDERATIONS

2.1. Elements of Robotic Capabilities

The primary goal of the test arenas is to provide reproducible measurements and tests of autonomous mobile robots. There are several elements that come together to create a fully autonomous mobile robot. Recognizing that there are going to be different levels of autonomy implemented in mobile robots, the arenas are designed to isolate the different capabilities that may be available on any particular robot. They are shown schematically in Fig. 1. For a more in-depth discussion of the design considerations for the arenas, see [2].

At the lowest level is the locomotion capability of the robot's physical platform. Although two of the three arenas provide some challenges for locomotion and require general agility of the robots, our emphasis (and that of the AAAI competitions) is on algorithms. So the arenas attempt to isolate and test the higher elements of robot autonomy and do not address locomotion directly.

The element just above the hardware implementation of locomotion and sensors is sensory perception. The robot has to sense what is in its environment in order to navigate, detect hazards, and identify goals (simulated victims and their locations). Sensor fusion is an important capability, as no single sensor will be able to identify or classify all aspects of the arenas. The simulated victims in the arenas are represented by a collection of different sensory signatures. They have shape and color characteristics that look like human figures and clothing. They have heat signatures representing body heat, along with motion and sound. The

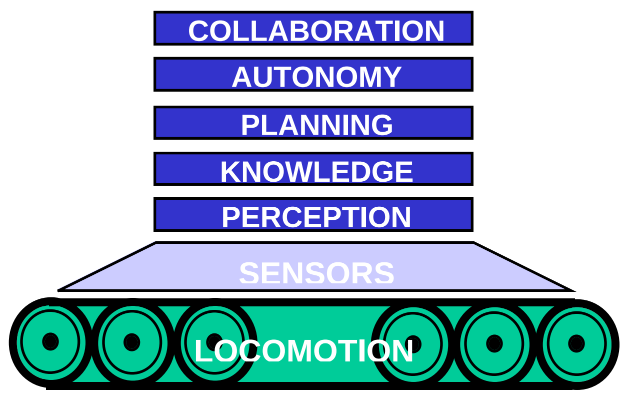


Figure 1: Constituent Elements of an Autonomous Mobile Robot

arenas are also designed to pose challenges to typical robot navigation sensors. For example, acoustic-absorbing materials confuse sonar sensors. Laser sensors have difficulty with shallow angles of incidence, smooth surfaces, and reflective materials. Highly regular striped wallpaper and other types of materials pose challenges to stereo vision algorithms. Compliant objects that may visually look like rigid obstacles require the robots to apply tactile sensors or other means of verifying that they can

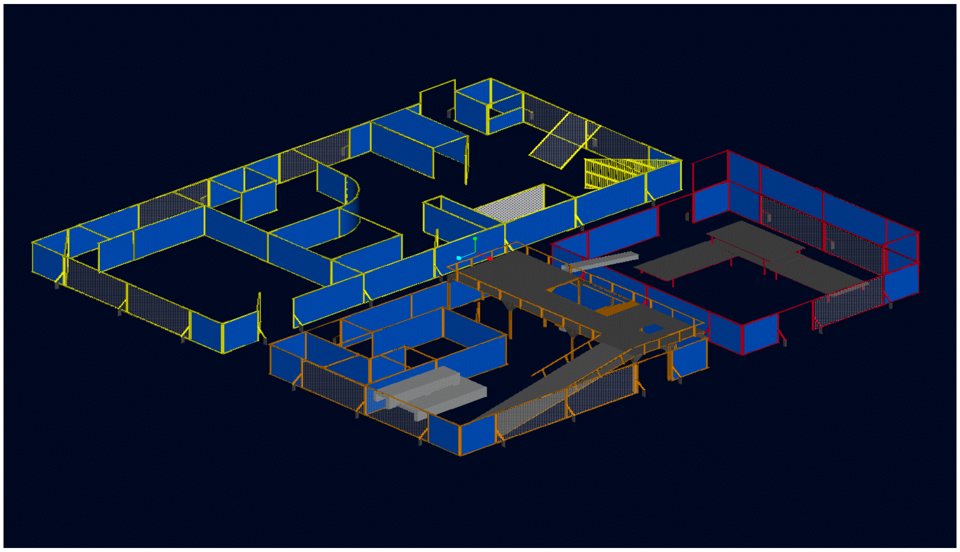
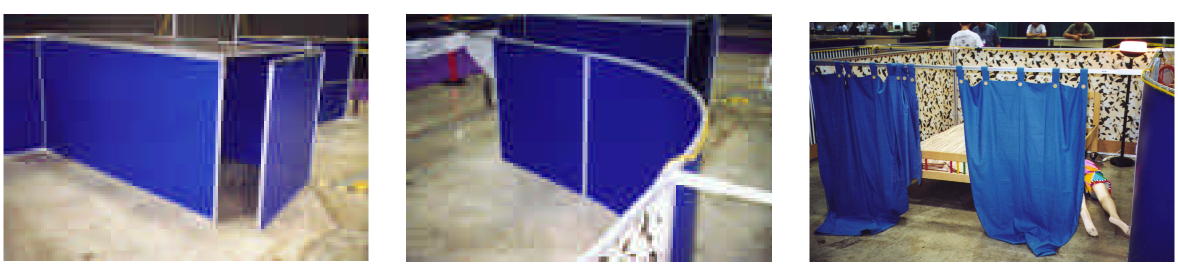


Figure 2: Model of the Reference Test Arenas for Autonomous Mobile Robots

indeed push them aside (e.g., open doors or curtains). Manipulation of rigid obstacles, such as closed doors or debris, provide more advanced challenges. Robot localization is another essential capability derived from sensing. Different flooring materials affect localization schemes based on wheel encoders. Additional cues from the environment need to be employed to help localize the robot in an effort to generate and maintain correct maps. Since the arenas represent collapsed structures and buildings, GPS is not considered to be available.

Knowledge representation is the next element. It encompasses the robot's ability to model the world, using both a priori information (such as might be needed to recognize certain objects in an environment) and newly acquired information (obtained through sensing the environment as it explores). In the mobile robot competitions for AAAI and RoboCupRescue, the robots are expected to communicate to humans the location of victims and hazards. Ideally, they would provide humans a map of the environment they have explored, with the victims' and hazards' locations marked. The environment that the robots operate in is three-dimensional, hence they should reason, and be able to map, in three dimensions. The arenas may change dynamically during a competition (as a building might further collapse while rescuers are searching for victims). Therefore the ability to create and use maps to find alternate routes is important.

The planning or behavior generation components of the robots build on the knowledge representation and the sensing components. The robots must be able to navigate around obstacles, make progress in their mission (that is to explore as much as possible of the arenas and find simulated victims), take into account time as a limited resource, and make time critical decisions and tradeoffs. The planner should make use of an internal map generated by the



a.) Darkened chamber with door **b.) Curved wall** **c.) Soft materials, victim under bed**

Figure 3: Features from the Yellow arena

robot and find alternate routes to exit the arenas that may be quicker or avoid areas that have become no longer traversible.
The overall autonomy of the robot is the next element to be evaluated. The robots must be designed to operate with humans. However, the level of interaction may vary significantly, depending on the robot's design and capabilities or on the circumstances. The intent is to allow for "mixed initiative" modes to limit human interaction, maximizing the effectiveness and efficiency of the collaboration between robot and humans. Robots may communicate back to humans to request decisions, but should provide the human with meaningful communication of the situation. Pure teleoperation is not a desirable mode for the robot's operation. The human should provide the robot with high level commands, such as "go to the room on the left" rather than joystick the robot in that direction.

The final element to be evaluated in the robot's overall capabilities is collaboration among teams of robots. One very rich area of research is in cooperative and collaborative robotics. Multiple robots, either heterogeneous or homogenous in design and capabilities should be able to more quickly explore the arenas and find the victims. The issues to be examined are how effectively they maximize coverage given multiple robots, whether redundancy is an advantage, and whether or how they communicate amongst themselves to assign responsibilities. Humans may make the decisions about assignments for each robot a priori, but that would not be as desirable as seeing the robots jointly decide how to attack the problem.

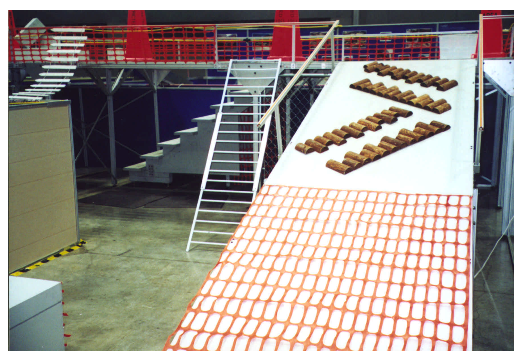
2.2. A CONTINUUM OF CHALLENGES

There are three separate Reference Test Arenas for Autonomous Mobile Robots, each labeled by a color denoting increasing difficulty. A schematic of all three arenas assembled together is shown in Figure 2.

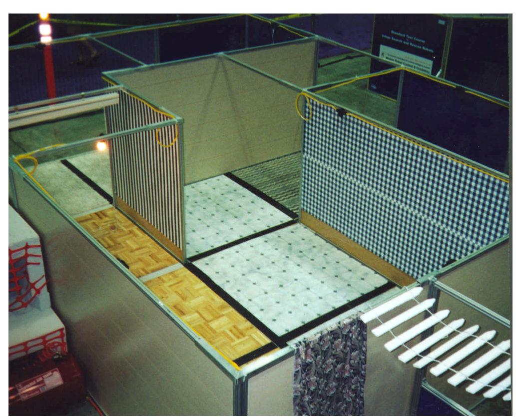
The Yellow arena is the easiest in terms of traversability. Researchers who may not have very agile robot platforms, yet want to test their sensing, mapping, or planning algorithms, can use the Yellow arena only. The arena consists of a planar maze. There are isolated sensor tests, based on obstacles or simulated victims. The arena is reconfigurable in real time, with doors that can be closed and blinds that can be raised or lowered. The reconfigurability provides challenges to the mapping and planning algorithms of the robots. A series of photographs of the Yellow arena features are shown in Fig. 3.

The Orange arena provides traversability challenges. Different types of flooring materials are present and there is a second story, reachable via ramp, stairs, and ladders. Holes in the second story floors requiring the perception, mapping, and planning capabilities of the robot be able to consider a three-dimensional world. The Orange arena is also reconfigurable in real time. Fig. 4 shows some features from the Orange arena.

The Red arena provides the least structure and the most challenges. It essentially represents a rubble pile (but is transportable). It is very difficult to traverse, with debris of various sorts throughout the arena. The debris is problematic for most robot locomotion mechanisms and includes rebar, gravel, plastic bags, and thin pipes. Simulated rubble resembling cinder blocks is strewn throughout. There are simulated pancaked floors (floors collapsed onto lower



a.) Ramp and other routes to 2nd story



b.) Different flooring materials and

Figure 4: Features of the Orange arena

Figure 5: Red Arena

floors) and leaning collapsed walls which can be triggered to cause secondary collapses. For example, the flooring in certain sections is unstable and will collapse if a robot attempts to surmount it. These features encourage robots toward a safer, more tactile approach toward negotiating the environment. A view of the Red arena is shown in Fig. 5.

3. THE 2001 COMPETITIONS

The NIST arenas made their debut at the AAAI-2000 Rescue Robot Competition [4][5]. Their second deployment was at the International Joint Conference on Artificial Intelligence (IJCAI) in 2001, where the RoboCupRescue and AAAI Robot Rescue competitions were jointly held.

In preparation for the second competition, a great deal of attention was paid to the development of scoring rules. The competition rules were designed to produce a final scoring distribution that defines clear winners. The focus of the competition is on intelligence; hence the scoring system favors solutions that demonstrate on-board autonomy, intelligent perception, world modeling, and planning. Fig. 6 shows the scoring formula.

Scoring is biased towards high quality interactions with humans, meaning that there is low-bandwidth, high content, infrequent communications to and from humans. The robots are expected to present human-understandable maps of their findings, highlighting the location of simulated victims. The scoring formula heavily favors multiple robots managed by a single operator. Improving the 1:1 ratio of operator to robot (teleoperation) is a key focus for these events. Simple teleoperative implementations, remotely using human perception for navigation and target acquisition, are not rewarded well in the scoring formula. The intent of these competitions is to push the state of the art toward autonomous solutions, while encouraging effective mixed-initiative modes of operation along the way.

Some disincentives were built into the scoring to discourage undesirable traits in the robots. For example, using simple redundancy of robots, while demonstrating no clear collaboration among the robots, implying the team could simply afford more robots, was discouraged. If the team could not demonstrate a cost-benefit advantage to having more robots (homogeneous or heterogeneous), their scoring suffered. In general, teams deploying multiple robots were penalized when their human-robot interface could not facilitate control of multiple robots by a single operator.

Other considerations in the design of the scoring were reflective of the course's design. "Gaming" of the arenas, that is, learning the course and its characteristics in order to "tune" the robots to perform well was obviously undesirable. Human level maps gained from operators closely scrutinizing the arena layout and simulated victim locations, and then teleoperating based on that knowledge, clearly undermines the intent of the competitions. But deterring that in the scoring was difficult. Since there were some fairly easy simulated victims to find, a minimum score was required to qualify for one of the place awards. The scoring formula also was designed to reflect the increasing difficulty of navigating and searching each progressively more challenging arena.

Six teams registered for the competition, but only four actually competed. No team scored enough points to qualify for either first, second, or third place awards. The two most successful teams earned "qualitative" awards for demonstrating very different capabilities.

$$\text{RobotRescueScore} = (\text{VictimsFound}\ (\text{NumberOfRobots} / (1+ \text{NumberOfOperators})^3)\ \text{AverageAccuracy}$$

$$\begin{aligned}\text{VictimsFound} = {} & (\text{VictimsFoundInYellow} / \text{VictimsPlacedInYellow})\ (\text{YellowVictimWeighting}) + \\ & (\text{VictimsFoundInOrange} / \text{VictimsPlacedInOrange})\ (\text{OrangeVictimWeighting}) + \\ & (\text{VictimsFoundInRed} / \text{VictimsPlacedInRed})\ (\text{RedVictimWeighting})\end{aligned}$$

$$\begin{aligned}[\ \text{YellowVictimWeighting} &= 0.50\ ] \\ [\ \text{OrangeVictimWeighting} &= 0.75\ ] \\ [\ \text{RedVictimWeighting} &= 1.00\ ]\end{aligned}$$

NumberOfRobots = Number of robots that find a unique victim

NumberOfOperators = Number of operators having touched the robot or are in the hot zone

AverageAccuracy = Average of the positional accuracy for each victim found

$[\ \text{VictimAccuracy} = (\text{IsVictimInVolume}) / (\text{StatedPositionalVolume})\ ]$

Figure 6: Scoring Formula at the 2001 RoboCup Rescue/AAAI Rescue Robot Competition

Swarthmore College (USA) demonstrated the most artificial intelligence capability, but only navigated within the easiest Yellow arena. The scoring formula required that the robots confined to the Yellow arena find all of the victims to earn the minimum score to qualify for a "place" award and be competitive with robots entering the other two more difficult arenas. They came close, finding all but one of the victims during one of their runs, falling just short of earning a "place" award. They received a "qualitative" for best artificial intelligence display.

Sharif University (Iran) demonstrated a more robust tracked robot, and even attempted to negotiate the Red arena. However, they had issues with their control strategy, bumping walls and obstacles frequently. They even triggered a secondary collapse of the pancaked flooring in the Red arena (an advanced obstacle). They resorted to identifying victims from outside the arena, but suffered from inherent inaccuracies in their approach. And they required too many human operators to manage their single robot, limiting their total score and keeping them from earning a "place" award. However, their effort was notable, and their robot mechanisms were well designed, so they earned a "qualitative" award for demonstrating the best hardware implementation. The experience will almost certainly allow them to improve their system for next year. Integration of more AI functionality should produce a very strong showing.

4. PROPOSED SCORING CHANGES

Given the experiences of two years of competitions within the Reference Test Arenas for Autonomous Mobile Robot, certain changes to the scoring seem reasonable. Note that these are the opinions of the authors and may or may not be reflected in the final rules for future mobile robot competitions.

The scoring formula should encourage robots to use a greater variety of sensors by awarding specific points for demonstrating superior sensory perception. This could be accomplished by awarding points for correctly identifying each sensor signature, or "sign of life," emitting from the simulated victims (form, heat, sound, motion). Since the simulated victims consist of various combinations of these sensor signatures, representing various states of consciousness and exposure, sensor fusion algorithms could deduce critical information regarding the state of the victim. This would allow more points to be scored per victim found, and would appropriately encourage the use of multiple sensors, along with sensory perception, sensory fusion, and error checking algorithms.

Some teams attempted to identify victims by looking through the clear windows on the perimeter of the arenas, thus avoiding the hazards within the harder arenas. The point values gained by identifying simulated victims from outside the course should be limited. The windows were placed to allow spectators visibility into the arenas, and to provide a

realistic obstacle for the robots. However, since no agility is required when the robot is outside of the arenas, the robot should not receive full credit for victims found in the harder Orange and Red arenas. The point values in such cases should be equivalent to finding victims in the Yellow arena.

Several behaviors exhibited by robots in the competitions should be discouraged through point deductions. Foremost should be point deductions for crushing, or inappropriately contacting, victims. Finding a victim (scoring points) and then hurting that victim should produce limited net gain in terms of scoring.

Causing damage to the arenas or certain obstacles through purposeful, or inadvertent, contact with the environment should also be discouraged with point deductions. If a robot triggers a secondary collapse of debris, the results could be catastrophic leading to further injuries or worse. These robots need to learn to be as deft as rescue personnel in their interactions with the environment, and should be penalized when they fail. There are a few typical voids in the arenas that can be destabilized and collapsed. Triggering these collapses should cause severe point deductions. Some lesser deduction should be tied to routine bumping of walls and other obstacles, demonstrating perception, planning, or control issues.

Also, teams which deploy more than one robot but sequentially teleoperate each one should be more effectively recognized in the scoring formula as maintaining a 1:1, operator:robot ratio, and not be lavishly rewarded as are multiple robot teams.

Lastly, maneuvering a robot based on human knowledge of the arena layouts or simulated victim placements essentially thwarts the spirit of the competition and should be discouraged. This is, of course, harder to implement in the scoring formula. However, focusing a larger percentage of the scoring potential toward autonomous activities (perception, control, planning, mapping, collaboration), while allowing some points for teleoperative techniques (identifying simulated human forms via remote video), the incentives would at least be in line with the goals of the competition.

5. FUTURE ACTIVITIES

NIST's Reference Test Arenas for Autonomous Mobile Robots will continue to be used to host the AAAI Rescue Robot Competitions in 2002. After two years of competitions, no robot team has demonstrated the minimum capabilities required to earn a "place" award. So it appears the research community has been challenged effectively. The RoboCupRescue competition has adopted these same arenas to host their competitions, and will use the same scoring formula developed for AAAI. Replicas of the arenas will be built for each RoboCupRescue event and left in the host country. This will result in the dissemination of the arenas worldwide, raise awareness of the needs and challenges for search and rescue robots, promote the competitions, and enable researchers to practice in the actual arenas throughout the year.

In order to further disseminate the arena's challenges and encourage progress in mobile robotics, NIST is developing virtual versions of the arenas. The effort is two-fold. Initially, sensor datasets obtained from within the arenas will be made available for download from the internet. This will permit researchers to process the data captured from sensors directly in the arenas and develop their algorithms without the need for problematic robot hardware. Data from a range-imaging sensor and from a color camera will be the first datasets available. A second, more ambitious, effort involves creating a simulated environment representing the arenas into which teams can plug their algorithms, receive simulated sensor data, and send actuation commands to navigate simulated robots. Further interaction with the research community is needed to design and develop this environment.

6. CONCLUSIONS

Tangible, realistic challenge problems can provide robot researchers with direction and help focus their efforts and collaborations. Reproducible, and widely known, challenges can help evolving fields by providing reference problems with measures of performance. Therefore, competitions, such as the AAAI Rescue Robot, RoboCupRescue, and others, can be valuable in spurring advancements in robotic capabilities. Thus far, the Reference Test Arenas for Autonomous Mobile Robots have been very

well received by the research community, and promise to provide a common set of reference challenges for the constituent elements of autonomous mobile robots. Their visibility in hosting competitions at AAAI, IJCAI, and other such events raises researcher's awareness of the types of challenges they must confront to be successful in the search and rescue domain. But the larger goal is to accelerate the advancement of mobile robotic capabilities through objective evaluation, collaboration, and the development of pertinent performance metrics, so that the capabilities that do emerge can be effectively applied to many other domains.

REFERENCES

[1] Blitch JG, " Artificial intelligence technologies for robot assisted urban search and rescue," EXPERT SYSTEMS WITH APPLICATIONS 11: (2) 109-124 1996.

[2] Jacoff, A., Messina, E., Evans, J., "A Standard Test Course for Urban Search and Rescue Robots," Proceedings of the 2000 Performance Metrics for Intelligent Systems Workshop, August 2000, Gaithersburg, MD.

[3] Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E., Matsubara, H., "RoboCup: A Challenge Problem for AI," AI Magazine 18(1): Spring 1997, 87-101.

[4] Murphy, R., Casper, J., Micire, M., Hyams, J., "Assessment of the NIST Standard Test Bed for Urban Search and Rescue Competitions," Proceedings of the 2000 Performance Metrics for Intelligent Systems Workshop, August 2000, Gaithersburg, MD.

[5] Schultz, A. (2001). "The 2000 AAAI Mobile Robot Competition and Exhibition," AI Magazine, 22(1), Spring 2001, AAAI, 67-72.